

• 新技术与新方法 •

多参数急救数据库建设及初步应用研究



王俊梅¹, 刘同波², 孙瑜尧³, 李沛尧⁴, 赵宇卓⁵, 张政波^{4,6,7}, 薛万国⁶, 黎檀实⁵, 曹德森^{4,7}

1. 北京航空航天大学生物与医学工程学院(北京 100191)
2. 解放军总医院计算机室(北京 100853)
3. 东南大学软件学院(江苏苏州 215123)
4. 解放军总医院医学工程保障中心(北京 100853)
5. 解放军总医院急诊科(北京 100853)
6. 解放军总医院医疗大数据中心(北京 100853)
7. 解放军总医院医疗器械研发与评价中心(北京 100853)

【摘要】 医疗大数据的分析利用离不开高质量的临床数据库,我国在急救数据库建设方面尚处于起步和探索阶段。本文介绍了多参数急救数据库的构建思路和关键技术,参考麻省理工学院(MIT)计算生理学实验室创建的重症监护医学信息数据库(MIMIC-III)的架构设计,并结合急诊业务流和信息流,设计了急诊数据整合模型,完成了高质量急救数据库的建设。该数据库目前涵盖了2015年5月至2017年10月共19 814名不同患者的22 941次抢救医疗数据,包含相对完整的生理、生化、治疗、检查、护理等信息,并基于该数据库开展了首届急救大数据 Datathon 活动,全国有13个队伍参赛。急救数据库的建设为国内临床数据库的构建和应用提供了参考,为科学研究、临床决策和改善医疗服务质量提供了有力的数据支撑,将进一步推动我国临床数据的二次分析利用工作。

【关键词】 急救数据库; 急危重症; 医疗数据库; 临床决策支持

Construction of multi-parameter emergency database and preliminary application research

WANG Junmei¹, LIU Tongbo², SUN Yuyao³, LI Peiyao⁴, ZHAO Yuzhuo⁵, ZHANG Zhengbo^{4, 6, 7}, XUE Wanguo⁶, LI Tanshi⁵, CAO Desen^{4, 7}

1. School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, P.R.China
2. Computer Department, Chinese PLA General Hospital, Beijing 100853, P.R.China
3. School of Software, Southeast University, Suzhou, Jiangsu 215123, P.R.China
4. Medical Engineering Support Center, Chinese PLA General Hospital, Beijing 100853, P.R.China
5. Emergency Department, Chinese PLA General Hospital, Beijing 100853, P.R.China
6. Medical Big Data Center, Chinese PLA General Hospital, Beijing 100853, P.R.China
7. Medical Device Research and Development and Evaluation Center, Chinese PLA General Hospital, Beijing 100853, P.R.China

Corresponding author: ZHANG Zhengbo, Email: zhengbozhang@126.com

【Abstract】 The analysis of big data in medical field cannot be isolated from the high quality clinical database, and the construction of first aid database in our country is still in the early stage of exploration. This paper introduces the idea and key technology of the construction of multi-parameter first aid database. By combining emergency business flow with information flow, an emergency data integration model was designed with reference to the architecture of the Medical Information Mart for Intensive Care III (MIMIC-III), created by Computational Physiology Laboratory of Massachusetts Institute of Technology (MIT), and a high-quality first-aid database was built. The database currently covers 22 941 medical records for 19 814 different patients from May 2015 to October 2017, including relatively complete information on physiology, biochemistry, treatment, examination, nursing, etc. And based on the database, the first First-Aid Big Data Datathon event, which 13 teams from all over the country participated in, was launched. The First-Aid database provides

DOI: 10.7507/1001-5515.201809032

基金项目: 国家自然科学基金(61471398); 北京市科委医药协同科技创新研究(Z181100001918023); 解放军总医院医疗大数据研发项目(2016MBD-027)

通信作者: 张政波, Email: zhengbozhang@126.com



a reference for the construction and application of clinical database in China. And it could provide powerful data support for scientific research, clinical decision making and the improvement of medical quality, which will further promote secondary analysis of clinical data in our country.

【Key words】 emergency database; acute and critical illness; medical database; clinical decision making

引言

目前医疗行业已经迅速实现了计算机化及数字化,大部分医疗机构都直接或间接地使用信息技术提供医疗服务。现代医疗信息系统能够产生并存储海量的数据,这些数据涉及详细的患者信息和临床诊疗过程,但是来自真实世界的患者数据却很少能被用于医学研究,推动医疗领域的创新和发展,其主要障碍在于研究人员很难获取和使用这些数据。若是能够整合并开放医疗机构内部以及不同医疗机构间的患者数据,就会吸引更多的、具有不同背景的研究人员,通过多学科协作共同解决目前临床医学中所面临的各类基础问题。

传统的医学研究收集数据的方法大多是依靠临床试验注册数据库或慢病管理系统,建立专病数据集。专病数据库中记录的数据往往局限于某一特定的研究任务,通常只针对某一特定的疾病,其数据或结论不能用于其他疾病,并且其数据经常是来自某一地区的特定人群,往往导致基于这些数据库的研究结论普适性较差。

第一个电子健康档案 (Electronic Health Records, EHR) 于 20 世纪 60 年代问世。EHR 的建立最初是为了归档整理患者的医疗记录,后来又增加了计费和质量改进的功能。之后 EHR 数据库不断发展,越来越全面,能够动态更新,并且信息相互关联。EHR 为建设更完整的医疗数据库提供了数据基础,进而为医疗大数据分析提供了机遇。它可以收集多维度的医疗数据,这使得临床医生、信息学家、数据工程师可以使用大型医疗信息数据库研究解答各种问题。在信息技术不断发展的今天,我们使用大数据挖掘理论和技术来分析和理解数据的能力都取得了巨大的进步,医疗数据挖掘分析在疾病风险评估、临床决策支持、疾病发展预测、合理用药指导、医疗管理和循证医学等领域发挥了重要作用^[1]。未来的临床研究将依赖于大数据技术来改善患者的诊疗,以期为患者提供更好的服务。

目前国外已有不少成熟的大型医疗数据库,可供全球研究人员开放使用。如:重症监护医学信息数据库 (Medical Information Mart for Intensive Care, MIMIC),包含了超过 6 万的重症加强护理

病房 (Intensive Care Unit, ICU) 诊疗数据^[2]; 还有飞利浦的 eICU 数据库,囊括了 2014 年到 2015 年间美国本土各地区十多万患者的 ICU 临床数据。这两个数据库都可以通过 <https://physionet.org/> 网站访问,通过授权后使用,支持研究人员开展各种临床研究,为医疗提供决策支持,如在大数据库中进行查询可以提升选择亚群的准确度,便于研究者对药品及医疗干预不良事件的监测,也可以帮助我们解决既往临床研究中不可测的问题^[3]。

在急诊救治过程中,会产生大量的动态医疗数据,其信息量丰富,价值密度大,建立大样本多参数急救数据库,可以为急救医学研究提供来自“真实世界”的数据,提高救治的时效性和准确性,进而发展基于数据驱动的辅助决策支持系统。目前已有许多基于急诊数据展开的研究,如美国俄勒冈州建立了全州范围内的急诊医疗服务数据库,增强了基于结局的急诊医疗服务研究,为提高医疗服务质量提供了有力的数据支持^[4]; Heldeweg 等^[5]使用研究型数据库 (REDCap) 中的急救数据,基于心电图 (electrocardiogram, ECG) 和心率变异性 (heart rate variability, HRV) 参数为急诊胸痛患者建立了性能优于心肌梗死溶栓治疗 (The Thrombolysis in Myocardial Infarction, TIMI) 危险评分的心血管危险分层模型,为临床决策提供了参考; Horng 等^[6]基于急诊数据,使用机器学习方法建立了可以识别脓毒血症感染患者的模型,为临床医生提供了良好的决策支持。

我国的临床数据库建设和应用尚处于起步阶段,目前存在的很多数据库是因课题需要而建设的小型数据集或者专科数据库^[7-11],这些数据库的数据来源较单一,共享性差,且质量良莠不齐^[12]; 国家人口健康科学数据共享平台提供了 200 多个科学数据集 (<http://www.ncmi.cn/>),但是这些数据颗粒度较低,信息不够完整。开源访问的高质量临床数据库和急救数据库建设在国内尚处于起步阶段。本课题组以解放军总医院急诊科为试点,以 MIMIC 数据库为参考,探索了我国急救数据库建设和应用思路。本文介绍了多参数急救数据库的构建思路和关键技术,以及基于该数据库开展的应用研究工作。

1 数据库架构分析与数据整合

1.1 MIMIC-III 数据库架构分析

自 2003 年以来,麻省理工学院 (Massachusetts Institute of Technology, MIT) 计算生理学实验室、贝斯以色列女执事医学中心和飞利浦医疗共同合作,在美国国立卫生研究所资助下,一起开发和维护了 MIMIC 数据库。该数据库收集了医院信息系统和 ICU 信息系统中存储的临床数据、床旁监护仪中的高精度波形数据以及来自社会保障总署的死亡信息。MIT 计算生理学实验室团队在 2015 年发布了重症监护医疗数据仓库 MIMIC-III^[13], 包含了 2001 年到 2012 年共 38 597 名去隐私化的成人入院患者数据, 共计 53 423 人次的 ICU 记录 (每名患者会有多次入院以及多次进出 ICU 的记录), 数据主要来源于内科重症监护室 (Medical Intensive Care Unit, MICU)、外科重症监护室 (Surgical Intensive Care Unit, SICU)、心内科监护室 (Cardiac/Coronary Care Unit, CCU) 和心脏外科监护室 (Cardiac Surgery Intensive Care Unit, CSRU) 这四个监护室。目前, 最新版 MIMIC-III v1.4 共包含 26 张数据

表, 主要分为四类: ① 定义和追踪患者, 如患者入院信息 (admissions)、患者入 ICU 信息 (icustays); ② ICU 内产生的患者数据, 如 ICU 所有记录观察结果 (chartevents)、患者入量 (inputevents_cv) 等; ③ 医院信息系统数据, 如实验室检查结果 (labevents)、诊断 (diagnoses_icd) 等; ④ 字典表, 如生化检查字典表 (d_labitems)、诊断字典表 (d_icd_diagnoses) 等。这些数据涵盖了患者进入 ICU、ICU 转换和出院的所有病程记录数据, 且数据经过日期 shift、格式转换等去隐私化处理, 最终形成该数据库, 其数据类型和整合流程如图 1 所示^[13]。

本文的急救数据库架构设计参考了 MIMIC-III v1.4 版本, 如图 2 所示, 不同类型信息存储在不同的数据表, 如 admissions 表存储了患者入院基本信息, 包括入出院时间、入院诊断和入院时基本状态等; patients 表存储了患者基本统计学信息, 包括性别、出生日期等; chartevents 表存储了患者生理生化等结果; d_labitems 以字典的形式存储生理生化等指标名称与 itemid 的映射关系。其中 subject_id 字段表示患者 id, 即对应唯一的患者, hadm_id 表示患者不同的入院记录 id, 患者入院信息、基本

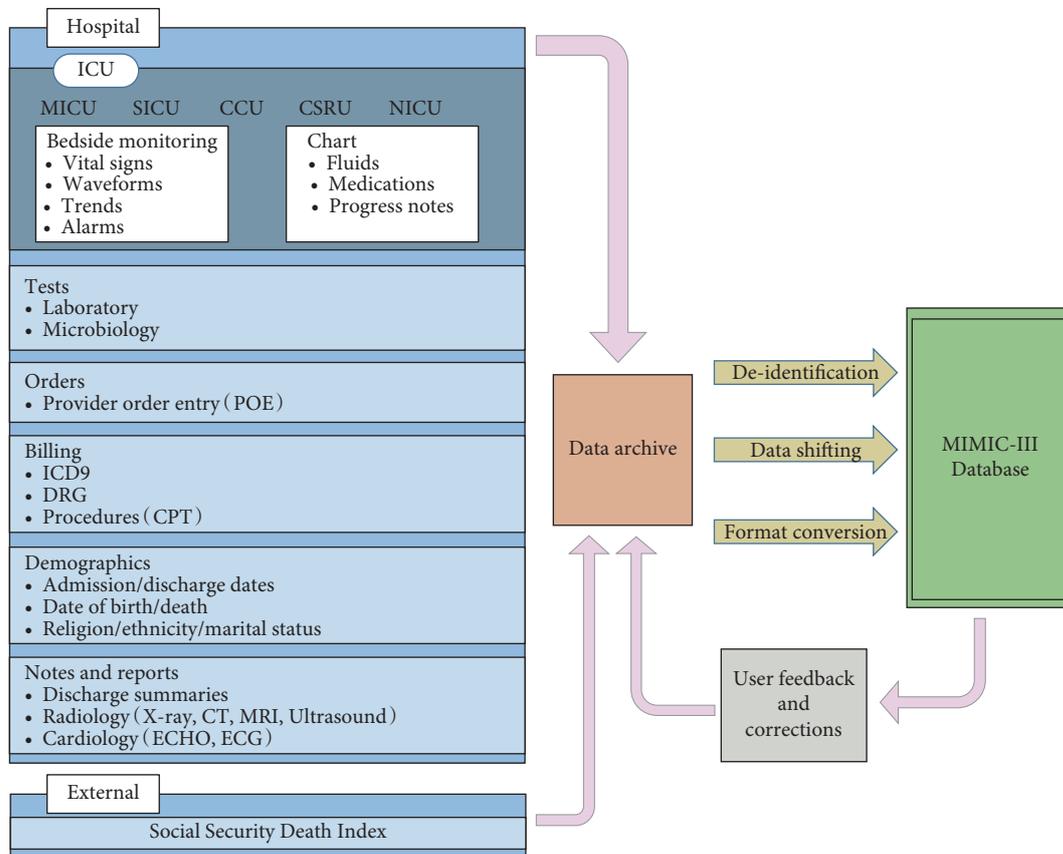


图 1 MIMIC-III 重症数据库概览
Fig.1 Overview of MIMIC-III Critical Database

统计学信息和其他临床诊疗信息通过这些 id 进行关联。另外,要获得患者的生理生化值需要先从 d_labitems 这样的字典表通过映射找到对应名称的 itemid, 并进一步查询获得结果。

MIMIC 数据库经过多个版本的更迭,在应用方面相当成熟,目前已有大量学者、研究员使用该数据库进行研究。同时,与数据库相关的问题会在 Github (<https://github.com/MIT-LCP/mimic-code/issues>) 上提出, MIT 团队会不断维护和完善数据库,使得数据库的架构更合理,易用性更高。目前该数据库在架构设计方面已经比较完善,因此可以作为全球急危重症数据库建设的参考标准。

1.2 急诊业务流程

为了建立合理的数据库架构,了解信息源,课

题组对急诊业务流程进行了深入调研,总结出急危重症患者从入院到出院的急诊业务流程,如图 3 所示。患者在急诊分诊台接受病情评估并被分诊到急诊相关科室,如急诊腹痛、急诊骨科、急诊妇产科等,如果患者有发热情况则会被分诊到发热门诊进行救治,若患者已排除患有传染病,可转入相应急诊科室就诊。病情比较危急的患者会进入急诊抢救室接受及时的救治,相应的,需要留院观察的患者会进入留观区,随着病情发展,患者可能进入专科、急诊病房、急诊重症监护室 (Emergency Intensive Care Unit, EICU) 以及转入其他医院继续接受治疗,或者宣布死亡。

要建设一个高质量的临床数据库,对急诊就诊业务流和信息流进行深入分析十分重要。急诊业

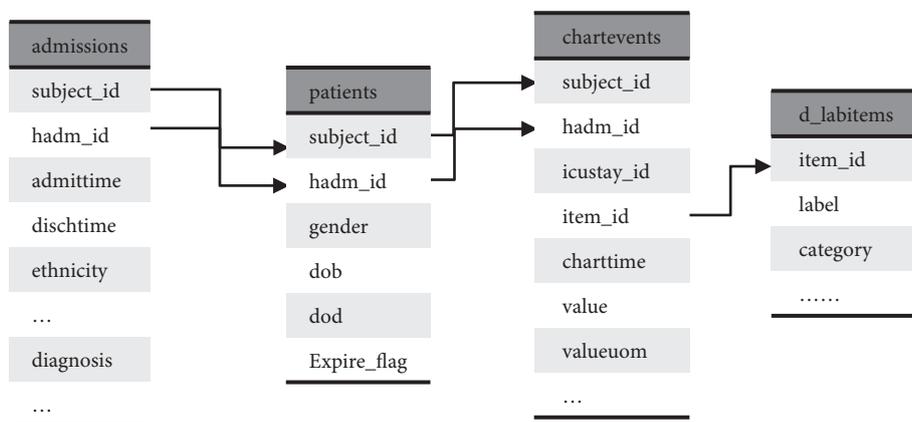


图 2 MIMIC-III 数据库表结构

Fig.2 Table structure of MIMIC-III Database

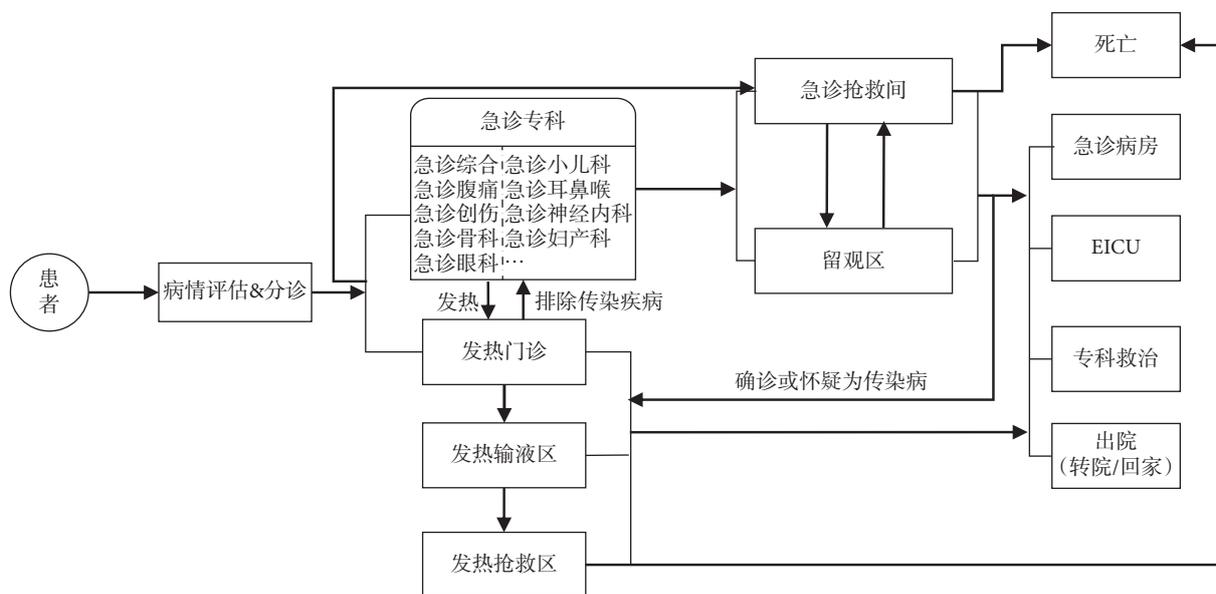


图 3 急诊患者就诊流程图

Fig.3 Treatment flow of emergency patients

务流和信息流涵盖了患者在就诊过程中产生的临床信息、信息类别以及信息之间的关联,有助于合理地设计数据库架构。患者在接受分诊时系统会记录其基本信息如就诊号、性别等,病情评估后患者被分诊到相应的科室或进入抢救室就诊,急诊抢救系统会记录患者在接受抢救的过程中的一系列数据,如床旁监护数据、生化检查、医嘱、护理和报告等数值和文本信息。

1.3 急救数据库数据整合

急救数据库数据主要来源于解放军总医院的医院信息系统(Hospital Information System, HIS)、急诊抢救系统和急诊专科诊室系统。急诊抢救系统和急诊专科诊室系统数据分别属于不同的医疗系统,从急诊专科诊室信息系统获得患者基本信息、分诊信息、就诊区域以及床旁监护数据;从急诊抢救系统获取患者就诊过程中的医嘱、用药和护理记录等信息。最后从医院 HIS 系统获得患者的生化检查数据,根据急诊患者的就诊流程,将患者从登记挂号到院内急救和转出这一过程中产生的一系列数据进行有效关联,数据整合流程如图 4 所示。患者可能有多次就诊记录,在该数据库中,患者的身份 ID 保持唯一,不同的就诊记录会分配有不同的就诊 ID 从而区别患者的多次就诊过程中产生的医疗信息。

1.4 数据预处理

在将患者基本信息及诊疗相关信息录入数据库前,需要对数据进行一定的处理,主要包括以下过程。

1.4.1 数据去隐私化 患者姓名、身份证号、家庭住址和家人信息等属于患者个人敏感信息,若这些信息发生丢失或泄露,不仅会给患者带来麻烦并造成经济损失,同时会给医院的声誉带来负面影响。因此将与患者真实信息相关的患者姓名和就诊的相关 ID (如 patientid、visitid 等) 进行去隐私化处理。患者相关的就诊 ID,一般由一位字母和一列数字组成,去识别化后的 ID 同样由字母和数字组成,因此首先获得所有患者原始 ID,对原始 ID 中首字母用其他字母如‘S’替代,然后以随机数的形式抽取数字部分并与某些数字做加减处理,产生的新数字与字母 S 组合产生新 ID,原始 ID 和新 ID 之间同时形成映射关系,因此可直接通过 ID 关联将新 ID 与患者就诊记录进行一致关联;患者姓名也借助算法产生新的名字,身份证和家庭住址等信息则去掉,不存储在该数据库中。对数据进行变换后,可避免直接或间接指向特定患者的信息出现在数据库中,同时保持原始数据相关特征和业务处理流程不变,即保持患者诊疗过程中所产生的其他一系列就诊相关信息不变。

1.4.2 数据清洗 解放军总医院急诊信息系统于 2014 年 11 月在急诊部门试运行,于 2015 年 5 月正式使用,而系统在试运行期间录入了一些测试数据,这些数据在正式运行之前并未清除,导致数据库中存在着歧义的数据:

(1)信息不一致。存在一部分数据,同一个患者 ID (patientid) 对应的患者基本信息不一致,例如

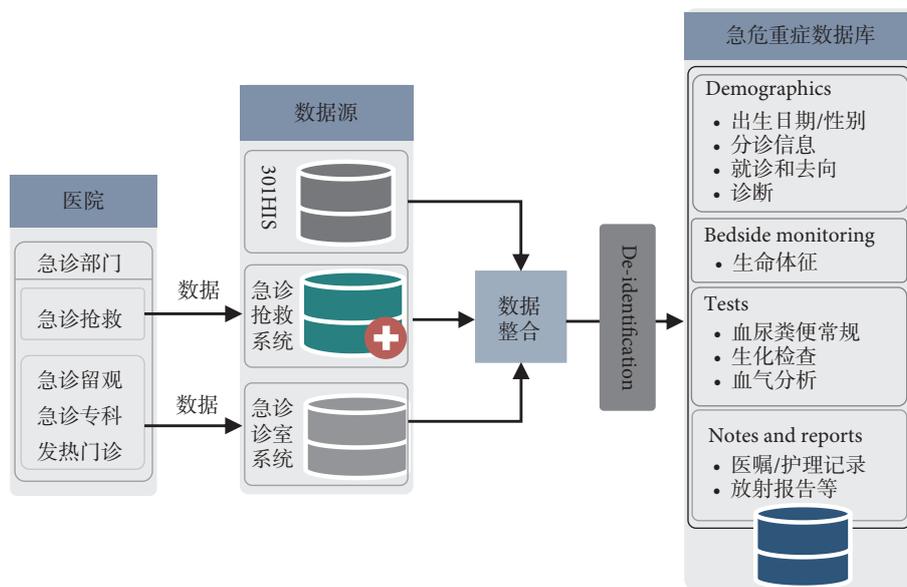


图 4 急救数据整合流程示意图

Fig.4 Schematic diagram of the emergency data integration process

patientID=1234321, 在患者多次就诊记录中, 存在出生日期不同或/和性别不一致的情况, 即同一患者 ID 两次或多次就诊时记录的性别(男/女)或者出生日期不同, 这样的信息会影响后续研究中的数据分析, 统计发现这样的数据约占总就诊量的 0.34%, 因此在数据中找出所有的患者 ID 相同但性别或出生日期不一致的数据, 然后删除以确保数据的一致性。

(2) 数据重复。由于护士操作或者系统卡顿原因, 数据库中存在对同一患者同一次就诊的重复分诊信息, 即在患者某一次就诊时, 除了分诊时间间隔相当短, 其他信息完全相同。在数据库建设过程中, 找出此类原因导致的多次重复分诊患者, 并去除患者的重复分诊记录。

(3) 无效数据。如患者的化验结果数据中存储了不可解读的错误信息, 例如化验结果“白细胞数”记录为“-----”, 在建库过程中通过正则表达式或者字符串匹配等方法, 找出所有存储结果无效的错误信息并予以去除。

2 数据库概述

急救数据库在国家自然科学基金面上项目(No.61471398)的资助下于 2017 年 7 月开始建设, 并在同年 12 月份测试完成。该数据库涵盖了解放军总医院 2015 年 5 月至 2017 年 10 月的急救患者诊疗数据, 由 19 814 名不同患者共 22 491 次的抢救医疗数据组成(很多患者会有多次入院以及多次进出抢救室的记录)。该数据库整合了患者在接受救治过程中产生的存储在多个医疗信息系统的临床相关数据, 共有 12 个数据表, 主要分为三类: (1) 定义、追踪患者, 如患者基本信息表(emg_patients)、

患者分诊信息(emg_triagePatients)等; (2) 抢救室内数据, 如床旁监护数据(emg_vitalSign)、医嘱信息(emg_medical_order)等; (3) 医院信息系统数据, 如辅助检查项目(exam_master、exam_report)、实验室检查(lab_test_master、lab_result)。表 1 给出了急救数据库中的具体的表结构和内容。

图 5 为某一患者入抢救室后的床旁监护数据, 该患者诊断为骨髓增生异常综合征和心房颤动, 床旁监护数据展示了数据库的高粒度特性和丰富的数据特征。表 2 给出了数据库中患者基本信息的统计。图 6 显示了急救数据库中患者分诊至各科室的统计分布, 其中急诊综合接收患者量最多, 急诊腹痛次之。表 3 描述了急救数据库中常见的急症, 其中消化系统急症、心血管系统急症和呼吸系统急症较为普遍。

3 数据库初步应用

数据库建设基本完成后, 我们对数据库的开放应用进行了探索。2018 年 3 月 16 日至 3 月 18 日, 解放军总医院急诊科联合清华大学、北京交通大学等国内 20 多家高校、科研院所, 并邀请来自国内外知名的临床和信息工程专家, 举办了第一届急救大数据 Datathon (<http://datathon.xomniam.com/>) 活动, 共同探索了多学科协作、开放共享的 Datathon 模式。本次比赛共有 13 个跨学科团队, 每支队伍由来自不同背景的临床医师和数据工程师组成, 首先提出研究问题, 制定研究方案, 利用解放军总医院急救数据库进行数据分析和建模, 并给出结论, 从而验证研究问题的合理性和有效性。整个实践过程均由专家老师指导, 临床医师与数据工程师分工协作, 综合运用各自的专业知识, 圆满地完成了数

表 1 急救数据库中的可用数据

Tab.1 Data available in the Emergency Database

数据表	描述
emg_patients	患者基本人口统计学信息(性别、出生日期、死亡日期)
emg_triagePatients	患者来院方式、分诊科室和分诊级别等信息
emg_visit	急诊抢救室患者信息(入出抢救室时间、诊断和去向信息)
emg_vitalSign	抢救室患者生命体征床旁监护信息(心率、呼吸、血压、氧饱和度、脉率和体温)
exam_master	患者辅助检查项目信息
exam_report	检查报告信息
lab_test_master	就诊患者化验检查申请记录
lab_result	患者化验结果信息
emg_chartevents	抢救室患者血气分析和心肌损伤五项检查信息
emg_medical_order	抢救室患者医嘱记录
emg_nursing_record	抢救室患者护理信息
emg_InOutRecord	抢救室患者出入量基本信息

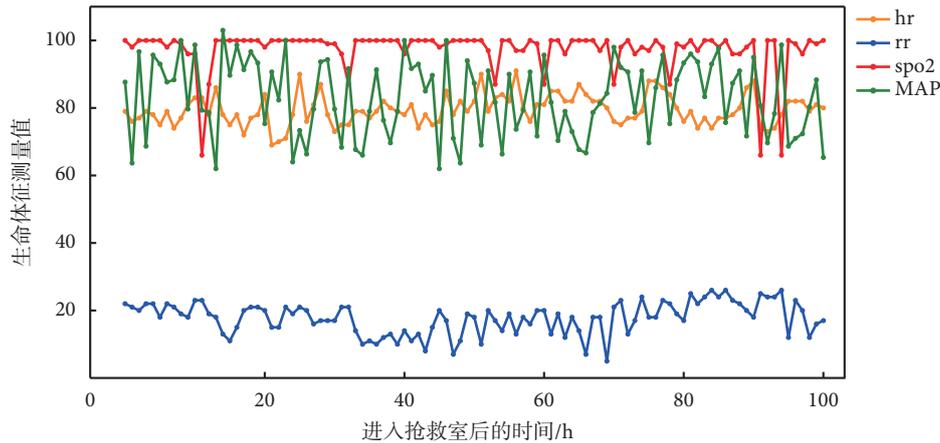


图5 某一患者在抢救室的床旁监护数据 hr: 心率; rr: 呼吸率; spo2: 血氧饱和度; MAP: 平均动脉压

Fig.5 Bedside monitoring data for a patient in the rescue room hr: heart rate; rr: respiratory rate; spo2: blood oxygen saturation; MAP: mean arterial pressure

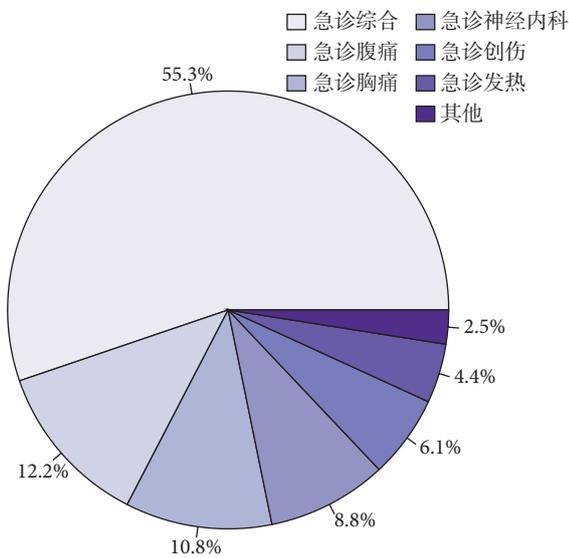


图6 急诊各专科急救患者就诊比例

Fig.6 Proportion of emergency patients in emergency departments

表2 急救数据库患者基本信息统计

Tab.2 Statistics of patients in the Emergency Database

数据表	统计
患者例数	19 814
就诊次数	22 491
年龄/岁[中位数(Q1~Q3)]	61.84 (46.55~76.98)
男性例数(百分比)	14 112 (62.7%)
急诊停留天数[中位数(Q1~Q3)]	0.50 (0.20~1.05)
急诊科内死亡例数(百分比)	545 (2.75%)

据实践活动。本次比赛中各参赛队伍研究题目如表4所示。

4 讨论与结论

本文详细介绍了多参数急救数据库的建设过

程,对数据库的结构和内容作了详细描述。数据库在架构设计方面参考了MIMIC-III^[13]数据库结构,内容上整合了来自解放军总医院急诊和部分专科系统的数据。在数据整合过程中,我们对患者临床数据进行了数据清洗以确保数据信息完整性和一致性并避免数据冗余。完成数据库建设后,我们进一步分析了数据库中存储数据的概况并对数据库的应用进行了初步探索。

MIMIC数据库经过10多年的持续优化和更新,数据库结构合理,易用性强,为全球急危重症数据库建设提供了参考标准。本文介绍的虽然是一个单中心(解放军总医院)临床数据库,但由于其参考了MIMIC数据库结构,结合了我院急诊救治的业务流和信息流,并对部分数据存储结构进行了改进,可以为建设我国高质量临床数据库提供有效的参考。如图7所示,床旁监护数据如心率(heartrate)、呼吸率(resprate)和氧饱和度(spo2)等不同生理指标会在同一时间被存储,图7右图展示了MIMIC-III数据库采用的行记录方式,即每一行记录一个时间对应的单个生理指标值,图7左图展示了本研究的数据库根据各个指标采样频率使用列存储的形式,每一行记录一个时间点对应的所有生命体征数据。一般研究者会同时使用所有的生命体征数据,本研究中的数据结构不仅便于使用者查询和使用数据,提高查询效率,也有助于节省存储空间,因此对国内急危重症数据库的建设具有一定的参考价值。

另一方面,数据库建设的目的在于分析利用,本文介绍的急救数据库也是国内首个可以向社会开放使用的高质量临床数据库,依托该数据库已经成功举办了第一届急救大数据和Datathon活动,

表 3 急救数据库患者急症信息

Tab.3 Emergency information of patients in the Emergency Database

急症	描述	例数 (百分比)
消化系统急症	急腹症、急性胰腺炎、消化道出血等	6 061 (27.0%)
心血管系统急症	急性冠脉综合征、心律失常、高血压急症、主动脉夹层、心跳骤停等	5 561 (24.7%)
呼吸系统急症	肺部感染、呼吸衰竭、支气管哮喘、张力性气胸、咯血、急性呼吸窘迫综合征等	5 200 (23.1%)
神经系统急症	脑梗死、脑出血、癫痫、颅内感染等	2 109 (9.4%)
外伤	车祸伤、坠落伤、锐器伤、电击伤等	1 376 (6.1%)
其他	中毒、电解质紊乱、贫血等	2 184 (9.7%)
总计		22 491

表 4 急救数据库 Datathon 参赛队伍研究内容

Tab.4 Datathon Team Research based on the Emergency Database

研究类型	研究题目
疾病预测预警模型研究	基于急救大数据的高乳酸血症短时长治疗效果预测模型研究 不同评分系统对急性重症胰腺炎预警作用比较 基于多主体筛选算法的消化道出血患者院内再出血治疗关键生理指标提取 动脉血乳酸联合 SIRS 在急危重症患者早期评估预警中价值研究 基于非支配排序筛选算法的急诊危重患者结局相关关键指标提取
疾病评估	呼吸衰竭患者结局相关因素分析 基于急救大数据的发热疾病早期诊断与病因预测 基于机器学习的急性冠状动脉综合征快速分诊评估 紧急救治危重患者相关危险因素分析
自然语言处理	影像报告自动语意分析
关联性分析	急性感染患者炎症因子水平与病原学关联研究 消化道出血风险与气象因素关系研究 基于 Apriori 算法的患者用药谱关联因素分析

有力推动了我国急救领域的临床数据资源建设和分析利用工作。目前该数据库依托于医疗大数据应用技术国家工程实验室和解放军总医院医疗大数据中心，向研究人员开放。经过审批后，研究人员可以在解放军总医院医疗大数据中心操作和使用该数据库，就感兴趣的临床问题开展研究。相信该工作将进一步推动我国临床数据库的建设和共享利用。

我国目前也存在很多医疗数据库^[14-16]，是针对

课题研究需要而建设的小型数据集或者专科数据库，这些数据库存储的是预先收集的患者数据，通常只包含某些特定的疾病，很少包含同一地区其他相关患者的数据，因此用这些数据库进行的研究普适性须进一步验证。而且由于国内很多 ICU 没有现代化的专科业务系统，无法获得更多的医疗设备数据，使得数据库中存储的生理数据缺失或较稀疏。急救患者有较高的恶化风险，如会发生感染、充血性心力衰竭^[17]、急性心肌梗死等严重病症，而

charttime	heartrate	reaprate	spo2	Spo2_pr
0:00:09	102	20	93	101
0:00:19	99	20	93	100
0:00:29	101	20	95	101
0:00:39	98	20	94	100
0:00:49	101	20	95	101
0:00:59	100	19	93	101
0:01:09	100	18	93	100
0:01:19	100	22	96	101
0:01:29	99	21	97	100
0:01:39	99	21	100	98

急救数据库

charttime	itemname	value	Valueuom
19:06	heartrate	115	bpm
19:16	heartrate	114	bpm
20:00	heartrate	113	bpm
21:00	heartrate	108	bpm
22:00	heartrate	110	bpm
23:00	heartrate	102	bpm
0:00	heartrate	108	bpm
1:00	heartrate	104	bpm
2:00	heartrate	93	bpm
3:00	heartrate	88	bpm

MIMIC-III 数据库

图 7 数据库床旁监护数据存储结构示例图

Fig.7 Storage structure of the database bedside monitoring data

生理数据在急危重症评估中有重要价值。因此要建设高质量的急危重症数据库,离不开现代化的专科业务系统和医疗设备物联网技术的支持。

我们的多参数急救数据库得益于急诊专科信息系统和医疗设备物联网的支持,可以采集到丰富的生理数据。与 MIMIC 数据库相比,该数据库还有一个特点是记录了更丰富完整的生化数据,如脑尿钠肽 (brain natriuretic peptide, BNP)、C-反应蛋白 (C-reactive protein, CRP) 和白细胞介素 6 (interleukin 6, IL-6) 等生化指标。其中 BNP 是急性冠脉综合征患者死亡的强关联指标,CRP 广泛用于急性感染性疾病的诊断和术后感染的监测,IL-6 能很好地评估急性胰腺炎的严重程度等,而 MIMIC 中这类重要的生化指标存在缺失或记录值较少。其中,在 MIMIC-III 数据库 61 532 次患者 ICU 就诊记录中,仅有 998 条 BNP 测量记录,即人均测量次数为 0.016,CRP 为 0.010,而 IL-6 无数据;在我们的急救数据库中,17 107 条不同就诊记录中,BNP 的人均测量次数为 1.367,CRP 为 0.861,IL-6 为 0.499。我们的数据库具有丰富完整的临床数据,可以支持研究者开展很多基于患者生理数据预测疾病发展趋势和疾病评估等的研究,有助于早期发现并干预有恶化风险的临床事件进而改善患者的预后,并节约医疗资源^[18-19],在以后的临床队列研究和改善医疗质量方面可起到重要作用,为临床研究提供数据支持进而改善急诊救治的效率和水平。

但是,在数据库建设过程中仍存在一些问題,该急救数据库仍需进一步完善,大致有以下几个方面:

(1) 数据库整合时,患者的实验室检查数据来源于院内 HIS 系统,而患者基本信息和就诊信息来自急诊诊室系统和急诊抢救系统,急诊信息系统和院内 HIS 系统之间的实验室检查数据关联较弱,是通过时间区间关联患者对应的数据,这导致患者的化验检查数据与患者的抢救记录(患者某一次入院记录中,存在多次时间间隔很短的抢救记录)无法直接匹配;

(2) 目前数据库仅包含患者在急诊部门的数据,而院内其他专科数据尚未录入数据库,无法追踪患者从急诊科转出至其他科室的医疗数据,无法获得患者在院内就诊的完整流程;

(3) 由于我们的数据库尚处于起步阶段,目前还未能与死亡登记数据库、国家医疗保险信息数据库等进行对接,所以与国际较成熟的医疗数据库相比存在一定不足。

接下来我们将从以下几个方面完善急救数据库:

(1) 进一步完善急诊信息系统,建立与院内 HIS 系统较强的关联,使得实验室检查数据能够与患者的某一次抢救室记录一一对应;

(2) 进一步整合院内其他科室的临床数据,使得患者从入院到出院过程中所有的临床数据被纳入数据库,形成完整的临床数据链;

(3) 完善数据库架构,为国内医疗数据库的建设提供参考。

随着院内更多高质量数据的整合,数据完整度不断提升,急救数据库将为大型临床研究的实施提供重要的平台,促进数据中潜藏信息的深度挖掘,进而为我国急危重症患者救治提供有力的数据支撑。

EHR 研究为研究者们提供了更多临床证据,研究者们可以使用大型临床数据库中的信息研究解答各种问题。EHR 为研究者们带来了研究模式的转变,利用 EHR 能帮助我们建立可持续学习的系统。与传统的随机对照试验 (randomized controlled trial, RCT) 相比,EHR 包含的变量更丰富,信息更详细完整,使用 EHR 可以完成许多 RCT 不能进行的研究,而且大大降低了成本和难度,是研究复杂的临床决策过程的理想工具。使用 EHR 还可以帮助临床医师精准地评估诊断结果,以及高效快捷地发现药品和其他医疗干预所导致的不良反应事件等^[5]。因此,临床医生和研究者们需要不断发掘海量 EHR 数据所蕴含的潜力,从而为临床提供更好的决策支持。

利益冲突声明: 本文全体作者均声明不存在利益冲突。

参考文献

- 1 Zhang Yue, Guo Shuli, Han Lina, *et al.* Application and exploration of big data mining in clinical medicine. *Chin Med J*, 2016, 129(6): 731-738.
- 2 Saeed M, Villarroel M, Reisner A T, *et al.* Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access intensive care unit database. *Crit Care Med*, 2011, 39(5): 952-960.
- 3 Ghassemi M, Celi L A, Stone D J. State of the art review: the data revolution in critical care. *Crit Care*, 2015, 19(1): 118.
- 4 Newgard C D, Zive D, Malveau S, *et al.* Developing a statewide emergency medical services database linked to hospital outcomes: a feasibility study. *Prehospital Emergency Care*, 2011, 15(3): 303-319.
- 5 Heldeweg M L A, Liu N, Koh Z X, *et al.* A novel cardiovascular risk stratification model incorporating ECG and heart rate variability for patients presenting to the emergency department with chest pain. *Critical Care*, 2016, 20(1): 179.

(下转第833页; Continued on Page 833)

